GENDER WAGE GAPS AND OAXACA DECOMPOSITION: TOOLS TO ACCOUNT FOR INDIRECT EFFECTS

Nick Deschacht

Department of Economics, KU Leuven

nick.deschacht@kuleuven.be

**2018 BELGIAN STATA USER GROUP MEETING**

1. Introduction

- The command `meddecomp` carries out a mediator decomposition, i.e. a Oaxaca decomposition in the presence of mediation effects.

- Oaxaca decomposition is often used to estimate to what extent gender differences in characteristics can explain the gender wage gap. In the example:

  - 327€ (45%) of the overall wage gap of 728€ can be explained by gender differences in characteristics.

  - 26% can be explained by the fact that men work more hours per week and 4% by the fact that men are more committed.

  - The sum of the contributions in the detailed decomposition equals the total explained part (45%).

	Euro	Pct.
Overall decomposition:		
Wage: men	2736	.
Wage: women	2007	.
Raw difference	728	100
explained	327	45
unexplained	401	55
Detailed decomposition:		
Working hours	188	26
Age	64	9
Degree	-66	-9
Industry (10 categories)	67	9
Organization size	47	6
Commitment	27	4
Sum of the contributions:		
Total		45
Observations: men	756	
Observations: women	609	

2. a) Oaxaca decomposition

The simple case of one explanatory variable:
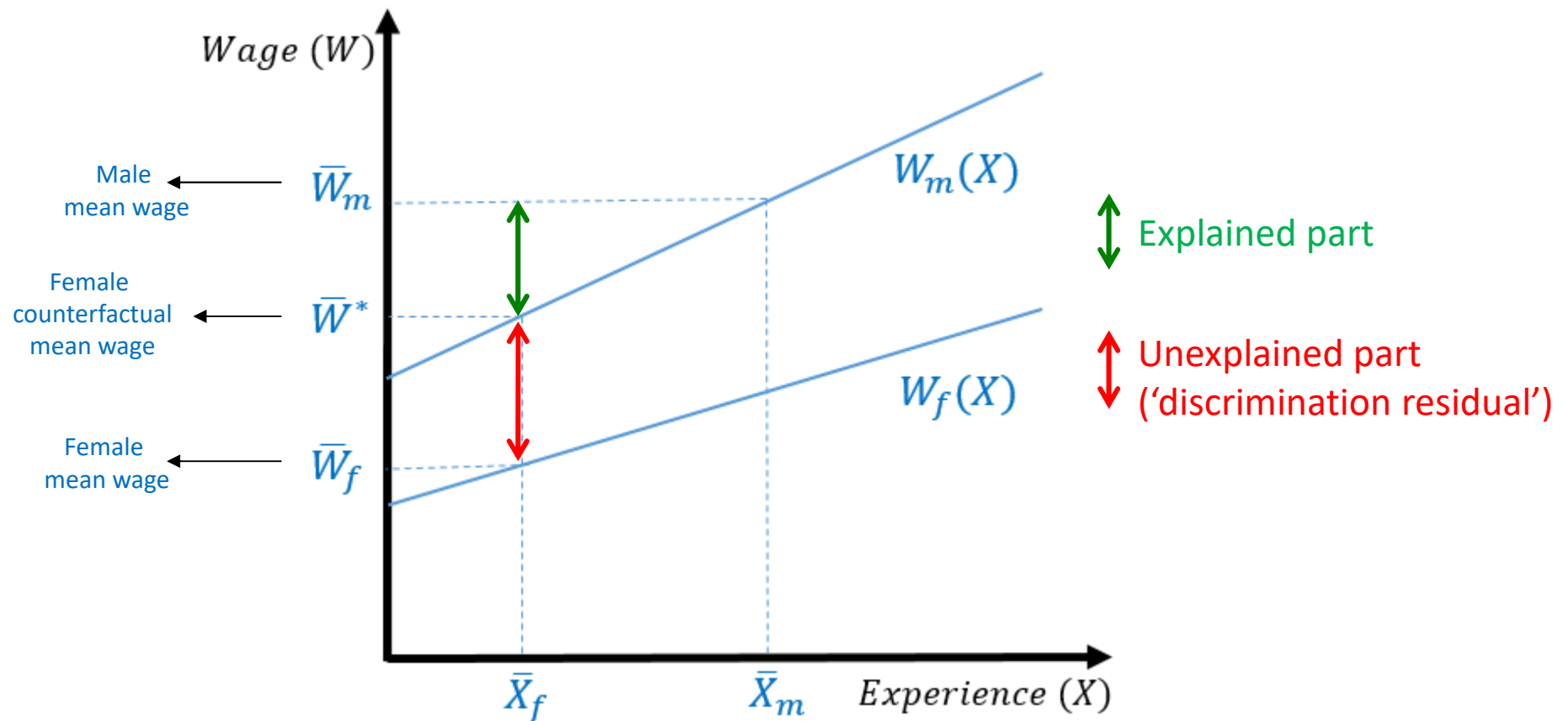
2. a) Oaxaca decomposition

- The contribution of a variable X to the gender wage gap is:

$$\left(\bar{X}_m - \bar{X}_f\right).b^*$$

 where b^* is the estimated effect of X on wages.

  E.g. if men have 20 years of experience, women only 18 years and b^* is 20 euro per year, then the part of the gender wage gap explained by the gender difference in experience is (20-18).20 = 40 euro.

- b^* can be estimated using the male sample only, the female sample or by pooling men and women. The choice depends on theoretical assumptions regarding discrimination, but pooling has the advantage that the coefficients are estimated using a larger sample (Neumark 1988).

- In Stata this decomposition can be done via the user-written command `oaxaca` (Jann, 2008). Extensions for nonlinear decomposition have been proposed by Yun (2005, built into `oaxaca`), Fairlie (`fairlie`: Jann, 2006) and Bartus (`gdecomp`).
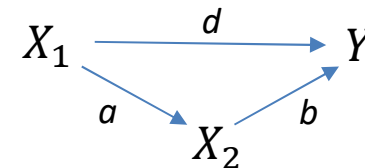
2. b) Mediation effects

- The effect of a variable X_1 on Y depends on the assumed underlying causal model. In the standard model, the effect of X_1 is the effect holding the other explanatory variables constant.

The standard model
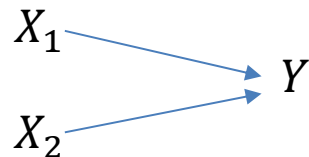
The mediator model

- In the mediator model, the total effect of X_1 is the sum of a direct effect (d) and an indirect effect ($a.b$) through an intermediate variable (the mediator).

  E.g. Catherine Hakim (2002) argues that preferences are more important than is often assumed because preferences not only affect wages directly, but also indirectly via working hours (part-time work).

- Stata estimates mediator effects via the commands `sem` and `gsem`.

3. The `meddecomp` command

```
Syntax

 meddecomp depvar indepvars , by(groupvar) med(mediator mediator_indepvars) [ options ]
```

- The command essentially performs a Oaxaca decomposition using the total effects (direct + indirect) obtained through a sem model.

  - `by` specifies the groups.

  - `med` specifies the mediator and the variable(s) that affect the mediator.

- `meddecomp` relies on the `oaxaca` command (Jann, 2008) for point estimations, standard errors and presentation of results.

- Example in Stata:

```
mydofile ✕
1
2     use deschacht_wagedata, clear
3
4     meddecomp wage hours commitment_high, by(gender_female) med(hours commitment_high)
5
```

3. The `meddecomp` command

	(1) Standard model		(2) Mediator model	
	Euro	Pct.	Euro	Pct.
Overall decomposition:				
Wage: men	2735.5	.	2735.5	.
Wage: women	2008.6	.	2008.6	.
Raw difference	726.9	100.0	726.9	100.0
explained	205.3	28.2	205.3	28.2
unexplained	521.6	71.8	521.6	71.8
Detailed decomposition:				
Working hours	156.6	21.5	153.5	21.1
Commitment	48.7	6.7	51.8	7.1
Sum of the contributions:				
Total		28.2		28.2
Observations: men	759		759	
Observations: women	611		611	

Compared to the standard model, the mediator decomposition results in the same overall decomposition (the total explained part is the same) but the detailed decomposition (the contribution of each explanatory variable) is adjusted to account for the indirect effects.

4. Command options and limitations

- `meddecomp` allows the aggregation of subsets of variables as in the oaxaca command.

```
mydofile* ×
1
2   meddecomp wage hours commitment_high (degree: degree_middle degree_high), by(gender_female) med(hours commitment_high )
3
```

- Categorical variables can be included using the prefix `xi`.

  Factor variables are not allowed because `meddecomp` relies on `oaxaca` and `sem`.

```
mydofile ×
1
2   xi: meddecomp wage hours commitment_high i.industry, by(gender_female) med(hours commitment_high)
3
```

- Example in Stata:

```
mydofile ×
1
2   xi: meddecomp wage hours age (commitment: commitment_middle commitment_high) ///
3   (degree: degree_middle degree_high) (industry: i.industry) (occupation: i.occupation), ///
4   by(gender_female) med(hours commitment_high)
5
```

4. Command options and limitations

```
Options
    linear        Adds decomposition results of a standard causal model (without
                  mediation).
    omega         Decomposition using pooled model coefficients excluding groupvar.
    adjust        Stores the estimated coefficients and adjusted means so that the
                  mediator decomposition can subsequently be done using oaxaca.
```

- The default decomposition uses the coefficients from a pooled model to which *groupvar* is added. Excluding *groupvar* (`omega`) is optional. The use of either male or female coefficients is not possible in this version.

- `adjust` allows the mediator decomposition to be done using `oaxaca` by adding the options `ref(meddecomp) x1($adjust)`.

  Example in Stata: results without standard errors (option `nose`)

```
mydofile ✕
1
2    meddecomp wage hours commitment_high, by(gender_female) med(hours commitment_high) adjust
3
4    oaxaca wage hours commitment_high, by(gender_female) ref(meddecomp) x1($adjust) nose
5
```

- `meddecomp` only allows one mediator and no more complex pathways than the simple mediator causal model.